Optimal Distributed Detection of Multiple Hypotheses Using Blind Algorithm

Optimal Distributed Detection of Multiple Hypotheses Using Blind Algorithm

By

Bin Liu

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Applied Science

McMaster University

by Bin Liu, October 2007

Master of Applied Science (2007) McMaster University
Electrical and Computer Engineering Hamilton, Ontario

TITLE: Optimal Distributed Detection of Multiple Hypotheses Using Blind Algorithm

AUTHOR: Bin Liu, M.A.Sc.
SUPERVISORS: Drs. Kon Max Wong, Aleksandar Jeremic

NUMBER OF PAGES: x, 51

Dedications:

To my family

Abstract

In a parallel distributed detection system each local detector makes a decision based on its own observations and transmits its local decision to a fusion center, where a global decision is made. Given fixed local decision rules, in order to design the optimal fusion rule, the fusion center needs to have perfect knowledge of the performance of the local detectors as well as the prior probabilities of the hypotheses. Such knowledge is not available in most practical cases. In this thesis, we propose a blind technique for the general distributed detection problem with multiple hypotheses.

We start by formulating the optimal M-ary fusion rule in the sense of minimizing the overall error probability when the local decision rules are fixed. The optimality can only be achieved if the prior probabilities of hypotheses and parameters describing the local detector performance are known.

Next, we propose a blind technique to estimate the parameters aforementioned as in most cases they are unknown. The occurrence numbers of possible decision combinations at all local detectors are multinomially distributed with occurrence probabilities being nonlinear functions of the prior probabilities of hypotheses and the parameters describing the performance of local detectors. We derive nonlinear Least Squares (LS) and Maximum Likelihood (ML) estimates of unknown parameters respectively. The ML estimator accounts for the known parametric form of the likelihood function of the local decision combinations, hence has a better estimation

accuracy.

Finally, we present the closed-form expression of the overall detection performance for both binary and M-ary distributed detection and show that the overall detection performance using estimated values of unknown parameters approaches quickly to that using their true values. We also investigate various impacts to the overall detection. The simulation results show that the blind algorithm proposed in this thesis provides an efficient way to solve distributed detection problems.

Acknowledgement

I would like to express my sincere appreciation to my supervisors, Dr. K. M. Wong and Dr. A. Jeremic for their guidance and encouragement throughout the course of this thesis, which have been inspiring my enthusiasm and creativity in research work. I also thank all of my colleagues here in the signal processing and communications group for their kind friendship and help. Last, I am deeply indebted to my family for their unselfish love and everlasting confidence in me along my life journey.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview of Distributed Detection

Hypothesis testing in distributed signal processing, referred to as distributed detection, is different in essence from that in classical multichannel scenarios. In the latter, usually referred to as centralized detection, observations from all channels are communicated to a central processor, where statistical inference on the hypotheses is conducted. However, many practical difficulties restrict the applicability of centralized detection, such as communication bandwidth, data transmission speed and computational complexity. In addition, observations collected from different channels could be incomparable and a decision on the hypothesis in question based on a mixture of observations may not be reliable [1]. In contrast to centralized detection, each local detector of a parallel distributed detection system preprocesses the observations it collects, makes a local decision and then transmits it to a fusion center where a global inference is made. Because the preprocessing occurs locally, the restrictions in centralized detection could be avoided. This kind of systems has been widely used in many real-world situations such as wireless communications, radar, control and biomedical engineering to enhance system survivability, to improve classification accuracy, and so on. In radar systems, a target identity is recognized according to

individual investigation of the incomparable target signatures by several radars. In biomedical engineering, results from different modalities are fused to draw a more accurate conclusion on the health condition of patients. These advantages are compromised by the reduced detection performance since some information contained in the original observations is lost due to the pre-procession at local detectors and the loss is not recoverable at the fusion center.

1.2 Literature Review

The literature of distributed detection is quite rich and continues to grow. The work of Tenney and Sandell [3] was deemed to be the pioneering publication in this area, where they derived the Bayesian formulation of the binary distributed detection. Under the assumption of statistical independence of the observations at local detectors conditioned on the hypothesis, the optimal local decision rules are likelihood ratio tests, with thresholds determined by a set of coupled nonlinear equations. However, the fusion rule was not taken into account to design or implicitly assumed to be known. Chair and Varshney [4], on the other hand, derived the optimal binary fusion rule in the Bayesian sense given the conditionally independent local detectors and fixed local decision rules. The global decision can be obtained by comparing the weighted sum of individual local decisions with a threshold. The weights are functions of the prior probabilities of the two hypotheses and the parameters describing the performance of local detectors. Hoballah and Varshney [11] attempted to optimize both local detectors and fusion center simultaneously and concluded that the person-by-person optimal necessary solution can be yielded by solving $N+2^N$ nonlinear coupled equations, where N is the number of local detectors. The computational complexity is prohibitively high in general and can be reduced when the observations at local detectors are independent. If the independence assumption of the observations does not hold, in general the problem is intractable because the necessary conditions for

optimality described by a set of coupled equations are extremely hard to solve. Zhu et al. [13] not only provided necessary conditions for optimum local decision rules under a given fusion rule when the observations of local detectors are dependent and proposed a discretized Gauss-Seidel iterative algorithm, but also proved in certain cases a fixed fusion rule can be used to achieve overall optimal performance. The distributed detection system can also be designed according to the Neyman-Pearson criterion. To facilitate the design, most researchers focus on the local decision rules or fusion rule individually. Similar to [3], Srinivasan [14] assumed the fusion rule is fixed and tried to optimize the local detectors. Thomopoulos et al. [15], on the other hand, carried out the optimization process only for the fusion center. The review papers [1, 2] contains more details on fundamental theory and advanced topics of distributed detection.

In this thesis, we focus on the design of optimal fusion rule in the Bayesian sense with fixed local decision rules, hence next we will only elaborate the literature related to this topic. It has been shown in [4] for binary distributed detection, the optimality can only be achieved if the prior probabilities and the probability of miss and false alarm of each local detector are known, which gives us the necessary weights for the optimal fusion. However, they are usually unknown in practice and need to be estimated at the fusion center using local decisions [6]. Naim and Kam proposed an adaptive algorithm to estimate the unknown parameters in [16], or the necessary weights may be estimated directly using reinforcement learning as suggested by Ansari et al. in [5]. Both algorithms are biased because they are derived under the assumption that the global decision at the fusion center is correct. Their attempt of removing the bias either greatly increases the computational complexity or results to worse convergence behavior of the algorithm. To bypass the problems aforementioned, Mirjalily et al. found the unknown parameters can be yielded by analytically solving a set of nonlinear equations involving the probabilities of different decision

combinations at all local detectors, which although are unknown in practical applications either, could be replaced by their corresponding empirical probabilities [6]. The resulting estimates of unknown parameters are asymptotically unbiased and substantially more reliable. All of the three algorithms deal with binary hypothesis.

The distributed detection process with multiple hypotheses, usually referred to as M-ary distributed detection, has recently attracted wide interest [7, 8, 9]. This is because a large number of practical problems such as sleep EEG discrimination and fault diagnosis of power systems consist of multiple hypotheses. The work in [4] has been extended to M-ary case [7]. Similar to binary distributed detection, it cannot be implemented when the prior probabilities and system parameters are unknown. One possible approach is to break the multiple hypotheses into binary decisions and utilize a hierarchical fusion process [8]. The fusion center creates a hierarchical partition of the M-ary local decision space using the probabilistic model of the observations. At each stage, two local decision sets are compared against each other. The final decision is made at the last stage of the decision tree. Another algorithm proposes to stipulate each local detector to making binary decisions only and, based on the binary decisions, an M-ary global decision is arrived at [9]. In [10], a suboptimal solution to the M-ary problem is suggested by applying the blind adaptive algorithm suited to binary hypothesis in [6].

In this thesis, we consider the general M-ary distributed detection problem with fixed local decision rules. The major contribution is two fold. One is that since the analytical solution of unknown parameters is intractable due to much more unknown parameters involved than the binary case, we take advantage of the known parametric distribution of local decision combinations to estimate the unknown parameters. The other is that we present the analytical expression of overall error probability and explore the effect of our blind algorithm on it.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows.

In Chapter 2, we first present the terminologies and definitions necessary for the work, including the general parallel distributed detection structure and the optimality criterion. We then introduce the local decision rules and formulate the optimal fusion rule by assuming the local decision rules are fixed. Since in reality the parameters required for the optimality are unknown, they have to be estimated using local decisions.

In Chapter 3, we discuss two parameter estimation algorithms, namely nonlinear Least Squares and Maximum Likelihood estimation to estimate the unknown parameters, then compare their performance with the biased Cramér-Rao bounds.

In Chapter 4, we derive analytically the overall detection performance. As long as the prior probabilities and parameters describing the performance of local detectors are known, the overall error probability can be computed accordingly. Since in real life we only have their estimates, the impact of estimation error to the overall detection performance is also addressed.

In Chapter 5, a series of studies on the performance of distributed detection system are conducted in a variety of scenarios. These investigations support the claims in the previous chapters and show the applicability of our proposed blind algorithm.

In Chapter 6, conclusions and future research directions for distributed detection system are presented.

Chapter 2

Background and Problem Statement

In this chapter, we introduce the background required for distributed detection and elaborate the problems to tackle in this thesis. In the beginning, the general structure and terminologies commonly used in the distributed detection system are presented. As mentioned, our goal is to design the optimal fusion rule when the local decision rules are fixed. Hence in the subsequent sections, we introduce the well known local decision rules and focus on the optimization of the fusion rule in the Bayesian sense. Finally, the blind adaptive algorithm for the binary distributed detection and the difficulty in extending it to the M-ary scenario are discussed, respectively.

2.1 Basic Terminologies

The parallel distributed detection system with N local detectors is illustrated in Fig. 2.1 and the frequently used terminologies are as follows:

- $\boldsymbol{z}_j$: observation vector at local detector D_j, $\boldsymbol{z}_j \in \boldsymbol{R}^{d_j}$, $1 \leq j \leq N$, where $\boldsymbol{R}^{d_j}$ is the observation space of local detector D_j, d_j is its dimension;
- u_j: decision at local detector D_j;

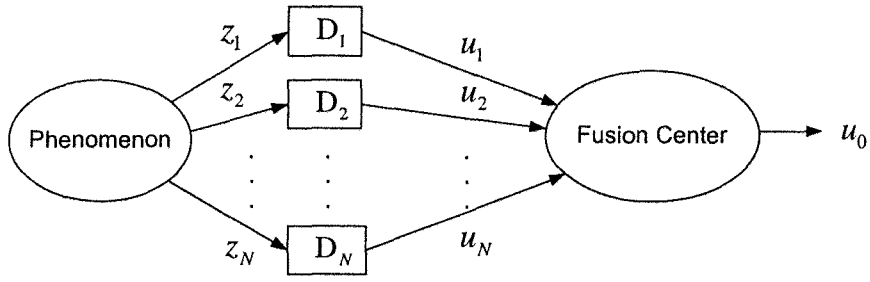

Figure 2.1: Parallel distributed detection system.

- u_0: global decision at the fusion center;
- $g_j(\boldsymbol{z}_j)$: decision rule of local detector D_j, which is a function of the observation;
- $g_0(u_1, \cdots, u_N)$: fusion rule, which is a function of all local decisions.

In general, the phenomenon changes from time to time, and at each time it could be one of the M possible hypotheses $\{H_0, H_1, \cdots, H_{M-1}\}$ with prior probabilities $P(H_i)$, $i = 0, \cdots, M-1$, respectively. For each time, the local detectors D_j, $j = 1, \cdots, N$ make a decision u_j individually according to their own observation $\boldsymbol{z}_j$. Given the unknown hypothesis H_i, the decision u_j is assumed to be conditionally independent of the decisions from other local detectors. The local detector D_j then sends u_j to the fusion center, where a global decision u_0 on the hypothesis is made based on a particular optimality criterion.

2.2 Local Decision Rule

Although the purpose of this thesis is to design optimal fusion rule when the local decision rules are fixed, for the completeness of the thesis we still discuss how the local decisions are made. In general, the observations of different hypotheses have different distribution. If the distribution function and prior probabilities are known, the local decisions can be made based on minimum error probability criterion, which

is equivalent to maximum a posteriori (MAP) criterion [17]. If the costs of making wrong decisions are given, minimum error probability criterion can be generalized to Bayesian criterion. In the case that there are only two hypotheses, it is also common to fix the value of probability of false alarm and maximize the probability of detection. This is called Neyman-Pearson criterion. No matter what criterion the local detectors stick to, we assume in this thesis that local decision rules are determined.

2.2.1 Example of Local Decision Rule

To make the explanation more clear, let us consider a local detector consisting of m spatially distributed sensors. The measurement of sensors located at the known positions $\{\boldsymbol{r}_i,\ 1 \leq i \leq m\}$ is generally modeled as

$$z(\boldsymbol{r}_i, t_j) = b(\boldsymbol{r}_i, t_j) + e(\boldsymbol{r}_i, t_j) \tag{2.1}$$

where $b(\boldsymbol{r}_i, t_j)$ is the true value of the measurement and $e(\boldsymbol{r}_i, t_j)$ is the corresponding measurement noise. The time samples are assumed to be taken at uniformly spaced time instants $\{t_j = jT_s,\ 1 \leq j \leq n\}$, where T_s is the sampling interval and n is the total number of time samples.

The linear source-measurement model is commonly used in sensor array community due to its simplicity. We suppose there are p sources, then stack all measurements into a vector. Let

$$\boldsymbol{z}(t_j) = [z(\boldsymbol{r}_1, t_j), \cdots, z(\boldsymbol{r}_m, t_j)]^T \tag{2.2}$$

and

$$\boldsymbol{z} = [\boldsymbol{z}^T(t_1), \cdots, \boldsymbol{z}^T(t_n)]^T \tag{2.3}$$

The measurement noise vector $\boldsymbol{e}$ can be obtained in a similar way. As a result, the source-to-measurement relationship illustrated in Eq. (2.1) can be written in a compact matrix form

$$\boldsymbol{z} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x} + \boldsymbol{e} \tag{2.4}$$

where $\boldsymbol{z}$ is mn-dimensional measurement vector, $\boldsymbol{A}(\boldsymbol{\theta})$ is a $mn \times p$ transfer matrix representing the relationship between the source and sensor measurement, $\boldsymbol{\theta}$ is a known parameter vector describing the source and medium property, and $\boldsymbol{x}$ represents the source intensity.

The transfer matrix $\boldsymbol{A}(\boldsymbol{\theta})$ is given by

$$\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\boldsymbol{\theta}_1), \cdots, \boldsymbol{a}(\boldsymbol{\theta}_p)] \tag{2.5}$$

The linear model is often seen in many practical situations [20, 21]. Next we will look at an example found in the literature of dipole source localization.

Example. It is usually believed that the mechanism of the EEG generation can be physically described as a set of current sources distributed over a certain region of the cortex. These distributed current sources can be further simplified as current dipoles in some situations where they are evoked in response to sensory stimuli such as auditory, visual, etc. The relationship between the primary potential on the scalp and the source, generally represented by the quasi-static approximation of Maxwell's equations, is reduced to

$$b(\boldsymbol{r}, \boldsymbol{\ell}, t) = \frac{1}{4\pi\sigma_0} \frac{\boldsymbol{r} - \boldsymbol{\ell}}{\| \boldsymbol{r} - \boldsymbol{\ell} \|^3} \boldsymbol{q}(t) \tag{2.6}$$

where σ_0 denotes the constant isotropic conductivity of the head, $\boldsymbol{r} = [r_x, r_y, r_z]^T$ is the electrode's location on the scalp, $\boldsymbol{\ell} = [\ell_x, \ell_y, \ell_z]^T$ and $\boldsymbol{q}(t) = [q_x(t), q_y(t), q_z(t)]^T$ are the location and moment of the dipole source embedded in the human brain, respectively [19]. The source location is assumed to be fixed, whereas the source moment, a measure of the strength of the neural signal in the three spatial directions, may vary with time. The scalp EEG received from a group of electrodes around the surface of human head is used to detect the existence of dipole sources or estimate the parameters describing them. The EEG measurements are composed of the primary

potential and measurement noise. Let us suppose m electrodes collect EEG measurements at n time samples. The EEG measurements are originated from p dipole sources. At the jth time instant, $1 \leq j \leq n$, the m-dimensional EEG measurement vector is

$$\boldsymbol{z}(t_j) = \boldsymbol{G}\boldsymbol{q}(t_j) + \boldsymbol{e}(t_j) \tag{2.7}$$

where $\boldsymbol{G} = [\mathcal{G}(\boldsymbol{\ell}_1), \mathcal{G}(\boldsymbol{\ell}_2), \cdots, \mathcal{G}(\boldsymbol{\ell}_p)]$ is an $m \times 3p$ gain matrix. Each submatrix $\mathcal{G}(\boldsymbol{\ell}_k)$, $k = 1, 2, \cdots, p$ denotes an $m \times 3$ gain matrix, relating the kth dipole source to the EEG measurements. It only depends on the source's location $\boldsymbol{\ell}_k = [\ell_{kx}, \ell_{ky}, \ell_{kz}]^T$ and all of the electrodes' location, and has the following structure according to Eq. (2.6)

$$\mathcal{G}(\boldsymbol{\ell}_k) = \frac{1}{4\pi\sigma_0} \begin{bmatrix} \frac{r_{1x}-\ell_{kx}}{{d_{k1}}^3} & \frac{r_{1y}-\ell_{ky}}{{d_{k1}}^3} & \frac{r_{1z}-\ell_{kz}}{{d_{k1}}^3} \\ \frac{r_{2x}-\ell_{kx}}{{d_{k2}}^3} & \frac{r_{2y}-\ell_{ky}}{{d_{k2}}^3} & \frac{r_{2z}-\ell_{kz}}{{d_{k2}}^3} \\ \vdots & \vdots & \vdots \\ \frac{r_{mx}-\ell_{kx}}{{d_{km}}^3} & \frac{r_{my}-\ell_{ky}}{{d_{km}}^3} & \frac{r_{mz}-\ell_{kz}}{{d_{km}}^3} \end{bmatrix} \tag{2.8}$$

where $\boldsymbol{r}_i = [r_{ix}, r_{iy}, r_{iz}]^T$ denotes the ith electrode's location, known on the surface of the head sphere, d_{ki} denotes the distance between the kth dipole and ith electrode, i.e., $d_{ki} = \| \boldsymbol{r}_i - \boldsymbol{\ell}_k \|$. $\boldsymbol{q}(t_j) = [\boldsymbol{q}_1^T(t_j), \cdots, \boldsymbol{q}_m^T(t_j)]^T$. $\boldsymbol{e}(t_j)$ is an m-dimensional measurement noise vector.

We then take into account all of the n time instants and write the resulting source-to-measurement relationship in a more compact form. Let

$$\boldsymbol{z} = [\boldsymbol{z}^T(t_1), \cdots, \boldsymbol{z}^T(t_n)]^T \tag{2.9}$$

and

$$\boldsymbol{q} = [\boldsymbol{q}^T(t_1), \cdots, \boldsymbol{q}^T(t_n)]^T \tag{2.10}$$

Therefore,

$$\boldsymbol{z} = (\boldsymbol{I} \otimes \boldsymbol{G})\boldsymbol{q} + \boldsymbol{e} \tag{2.11}$$

where "$\otimes$" represents the Kronecker product, $\boldsymbol{I}$ is an $n \times n$ identity matrix, $\boldsymbol{e}$ is the corresponding noise vector. In the case of fixed orientation moments, the moments can be decomposed into a unit-norm orientation vector and a scalar intensity. As a consequence, the signal part of the model can be further simplified as the product of a new gain matrix and source intensity vector [20].

After considering a specific example, let us move back to general linear source-to-measurement model describing in Eq. (2.4). Obviously, the mn-dimensional vector $\boldsymbol{a}(\boldsymbol{\theta}_k)$, $k = 1, \cdots, p$ gives the true sensor measurement about the kth source with unit intensity. For example, its $(m(j-1)+i)$th component is the ith sensor measurement about the kth unit intensity source at the jth time sample without any measurement noise. It is in general assumed that the measurement noise is spatially and temporally uncorrelated and Gaussian distributed with zero mean and unknown variance σ^2.

The detection of source can be formulated as a hypothesis testing problem with two hypotheses:

H_0 : The source is absent and only noise exists

H_1 : The source is present

This formulation in essence tests between $x_k > 0$ for some k and $x_k = 0$ for all k. Under the assumption of Gaussian distributed measurement noise, the detection problem is actually a test between two probability distributions, i.e.,

$$H_0 : \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

$$H_1 : \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \sigma^2 \boldsymbol{I}), \text{ where } x_k \geq 0 \text{ for } k = 1, \cdots, p$$

We then apply the likelihood ratio test detector to the hypotheses. In order to do so, all parameters describing the probability distribution of the measurements have to be known. However, unknown deterministic parameters exist in the probability

density function (pdf). To overcome this problem, the generalized likelihood ratio test (GLRT) is commonly used [17]. The unknown parameters under different hypotheses need to be estimated and their estimates are then substituted into LRT as if they were true values of unknown parameters. The estimates are usually obtained using maximum likelihood estimation. Let $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ represent unknown parameters for the pdf under H_0 and H_1, respectively, thus the GLRT is written as

$$\Lambda_g(\boldsymbol{z}) = \frac{\max_{\mu_1} p(\boldsymbol{z}|\boldsymbol{\mu}_1)}{\max_{\mu_0} p(\boldsymbol{z}|\boldsymbol{\mu}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.12}$$

In other words, we find the estimate of $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ maximizing the likelihood functions under their corresponding hypotheses and then use them in the LRT.

The GLRT in our case can be written as

$$\Lambda_g(\boldsymbol{z}) = \frac{\max_{x_k>0,\sigma^2>0} p(\boldsymbol{z}|x_k,\sigma^2)}{\max_{\sigma^2>0} p(\boldsymbol{z}|\sigma^2)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.13}$$

where

$$p(\boldsymbol{z}|H_1) = \frac{1}{(2\pi\sigma^2)^{\frac{mn}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{z}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{z}-\boldsymbol{A}\boldsymbol{x})\right\} \tag{2.14}$$

and

$$p(\boldsymbol{z}|H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{mn}{2}}} \exp\left(-\frac{\| \boldsymbol{z} \|^2}{2\sigma^2}\right) \tag{2.15}$$

It is well known that if there is no constraints about source intensity $\boldsymbol{x}$, the ML estimate of $\boldsymbol{x}$ under H_1 is

$$\begin{aligned} \tilde{\boldsymbol{x}} &= [\tilde{x}_1, \cdots, \tilde{x}_p]^T \\ &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{z} \end{aligned} \tag{2.16}$$

Given the constraint of non-negativity of the source intensity, the estimate should be modified as [29]

$$\hat{\boldsymbol{x}} = [\hat{x}_1, \cdots, \hat{x}_p]^T \tag{2.17}$$

where

$$\hat{x}_k = \begin{cases} 0, & \text{if } \tilde{x}_k \leq 0 \\ \boldsymbol{\alpha}_k^T \boldsymbol{A}_l^T \boldsymbol{z}, & \text{if } \tilde{x}_k > 0 \end{cases} \tag{2.18}$$

Here $\boldsymbol{\alpha}_k$ is the kth column of matrix $(\boldsymbol{A}_l^T\boldsymbol{A}_l)^{-1}$, l is the number of positive elements of $\tilde{\boldsymbol{x}}$, $\boldsymbol{A}_l$ is the $mn \times l$ submatrix of $\boldsymbol{A}$ whose columns correspond to the positive estimates and $\tilde{x}_k > 0$.

Once the estimate of $\boldsymbol{x}$ is obtained, we can estimate σ^2 under H_1 by substituting it into Eq. (2.14) and maximizing the resulting function with respect to σ^2. Its estimate under H_0 is achieved by directly maximizing Eq. (2.15). Hence,

$$\hat{\sigma^2} = \begin{cases} \frac{1}{mn}(\boldsymbol{z} - \boldsymbol{A}\hat{\boldsymbol{x}})^T(\boldsymbol{z} - \boldsymbol{A}\hat{\boldsymbol{x}}), & \text{when } H_1 \text{ is true} \\ \frac{1}{mn}\boldsymbol{z}^T\boldsymbol{z}, & \text{otherwise} \end{cases} \tag{2.19}$$

The GLRT can be rewritten by substituting all of the estimates into Eq. (2.13)

$$\Lambda_g = \left(\frac{\boldsymbol{z}^T\boldsymbol{z}}{\boldsymbol{z}^T\boldsymbol{z} - \boldsymbol{z}^T\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}}\right)^{\frac{mn}{2}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.20}$$

where

$$\boldsymbol{P}_{\boldsymbol{A}_l} = \boldsymbol{A}_l(\boldsymbol{A}_l^T\boldsymbol{A}_l)^{-1}\boldsymbol{A}_l^T \tag{2.21}$$

is the projection matrix of $\boldsymbol{A}_l$ onto the space spanned by the columns of $\boldsymbol{A}_l$. Let $\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}_l} = \boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}$. To simplify the generalized likelihood ratio (GLR), we then apply monotonic transformation $\Lambda_g^{2/mn} - 1$. Hence the GLR can be redefined to be

$$\text{GLR} = \frac{\boldsymbol{z}^T\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}}{\boldsymbol{z}^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}\boldsymbol{z}} \tag{2.22}$$

The decision rule is therefore

$$\text{GLR} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{2.23}$$

where $\tau = \eta^{2/mn} - 1$. The numerator and denominator can be viewed as the squared norm of Gaussian random vector $\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}$ and $\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}\boldsymbol{z}$, respectively. We known that under H_0,

$$\begin{aligned} E(\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}) &= \boldsymbol{P}_{\boldsymbol{A}_l}E(\boldsymbol{z}\boldsymbol{z}^T)\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \\ &= \sigma^2\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \\ &= 0 \end{aligned} \tag{2.24}$$

and

$$E(\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z})E(\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}\boldsymbol{z})^T = 0 \tag{2.25}$$

This shows that the two random vectors are uncorrelated. Furthermore, since they are Gaussian distributed, they are independent of each other. Therefore, the numerator and denominator are χ^2 distributed with l and $mn-l$ degree of freedom, respectively [28]. This means that the resulting division (GLR) follows central F distribution with degree of freedom l and $mn-l$. Similarly, under H_1,

$$E(\boldsymbol{z}\boldsymbol{z}^T) = \sigma^2\boldsymbol{I} + \boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T \tag{2.26}$$

hence,

$$\begin{aligned} E(\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}) &= \boldsymbol{P}_{\boldsymbol{A}_l}E(\boldsymbol{z}\boldsymbol{z}^T)\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \\ &= \boldsymbol{P}_{\boldsymbol{A}_l}(\sigma^2\boldsymbol{I} + \boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T)\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \\ &= \boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \end{aligned} \tag{2.27}$$

Furthermore,

$$\begin{aligned} E(\boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z})E(\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp}\boldsymbol{z})^T &= \boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{z}E(\boldsymbol{z})E(\boldsymbol{z})^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \\ &= \boldsymbol{P}_{\boldsymbol{A}_l}\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{P}_{\boldsymbol{A}_l}^{\perp} \end{aligned} \tag{2.28}$$

Again, this shows the two random vectors are uncorrelated and hence independent as a result of their Gaussian distribution. The numerator and denominator of the GLR are noncentral χ^2 distributed. Hence the GLR is noncentral F distributed with l and $mn-l$ degree of freedom and non-centrality factor

$$\lambda = \frac{\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}}{\sigma^2} \tag{2.29}$$

Let $F_{l,mn-l}(\lambda)$ denote the noncentral F distribution function having l and $mn-l$ degree of freedom. The variable λ is the non-centrality factor. Obviously, for central F distribution, $\lambda = 0$. In essence, the GLR is random variable and can take on several

values depending on l, i.e.,

$$\text{GLR} = \begin{cases} 1, & \text{if } l = 0 \\ F_{l,mn-l}(\lambda), & \text{if } 1 \leq l \leq n \end{cases} \tag{2.30}$$

By definition, the probability false alarm is

$$\begin{aligned} P^f &= P(\text{GLR} > \tau | H_0) \\ &= P[F_{l,mn-l}(0) > \tau] \\ &= \sum_{l=0}^{n} P[F_{l,mn-l}(0) > \tau | L = l] P(L = l) \end{aligned} \tag{2.31}$$

and the probability of miss is

$$\begin{aligned} P^m &= P(\text{GLR} < \tau | H_1) \\ &= P(F_{l,mn-l}(\lambda) < \tau) \\ &= P[F_{l,mn-l}(\lambda) < \tau | L = l] P(L = l) \end{aligned} \tag{2.32}$$

where L is a random variable representing the number of positive components in $\tilde{x}$. The derivation of the probability $P(L = l)$ is given in [21].

Note that in the aforementioned example, the local detector consists of m sensors. In some special cases, the local detector is simply a sensor, thus the noncentral F distribution function has l and $n - l$ degree of freedom.

2.3 Optimal Fusion Rule

In this section, we formulate both binary and M-ary fusion rule for fixed local decision rules.

2.3.1 Binary Hypothesis

Let us first consider a hypothesis testing problem with only two hypotheses H_0 and H_1. The prior probabilities are denoted by $P(H_1) = P_1$ and $P(H_0) = P_0$. As explained in

Section 2.1, each local detector employs a decision rule $g_j(\boldsymbol{z}_j)$ and concludes that

$$u_j = \begin{cases} 1, & \text{if } H_1 \text{ is claimed} \\ 0, & \text{if } H_0 \text{ is claimed} \end{cases} \tag{2.33}$$

Let $\boldsymbol{u} = (u_1, \cdots, u_N)$, we want the error probability defined as

$$P_e = P_1 P(u_0 = 0|H_1 \text{ is true}) + P_0 P(u_0 = 1|H_0 \text{ is true}) \tag{2.34}$$

to be as small as possible. According to binary detection theory in [17], the optimality criterion in the sense of minimum error probability can be written as

$$\frac{P(\boldsymbol{u}|H_1)}{P(\boldsymbol{u}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} \tag{2.35}$$

We rewrite Eq. (2.35) using Bayes' rule as

$$\frac{P(H_1|\boldsymbol{u})}{P(H_0|\boldsymbol{u})} \underset{H_0}{\overset{H_1}{\gtrless}} 1 \tag{2.36}$$

i.e.,

$$\log \frac{P(H_1|\boldsymbol{u})}{P(H_0|\boldsymbol{u})} \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{2.37}$$

Furthermore,

$$\begin{aligned} P(H_1|\boldsymbol{u}) &= \frac{P_1 P(\boldsymbol{u}|H_1)}{P(\boldsymbol{u})} && (2.38) \\ &= \frac{P_1}{P(\boldsymbol{u})} \prod_{S+} P(u_j = 1|H_1) \prod_{S-} P(u_j = 0|H_1) \\ &= \frac{P_1}{P(\boldsymbol{u})} \prod_{S+} (1 - P_j^m) \prod_{S-} P_j^m && (2.39) \end{aligned}$$

where $S+$ is the set of all indices j such that $u_j = 1$ and $S-$ is the set of all indices j such that $u_j = 0$. Similarly,

$$P(H_0|\boldsymbol{u}) = \frac{P_0}{P(\boldsymbol{u})} \prod_{S+} P_j^f \prod_{S-} (1 - P_j^f) \tag{2.40}$$

Substituting Eq. (2.38) and (2.40) into (2.37), we then have

$$\log \frac{P(H_1|\boldsymbol{u})}{P(H_0|\boldsymbol{u})} = \log \frac{P_1}{P_0} + \sum_{S+} \log \frac{1 - P_j^m}{P_j^f} + \sum_{S-} \log \frac{P_j^m}{1 - P_j^f} \tag{2.41}$$

Therefore we can express the optimal binary fusion rule as

$$u_0 = \begin{cases} 1, & \text{if } w_0 + \sum_{j=1}^{N} w_j > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.42}$$

where

$$w_0 = \log \frac{P_1}{P_0} \tag{2.43}$$

and

$$w_j = \begin{cases} \log \frac{1-P_j^m}{P_j^f}, & \text{if } u_j = 1 \\ \log \frac{P_j^m}{1-P_j^f}, & \text{otherwise} \end{cases} \tag{2.44}$$

The prior probabilities and the probabilities of miss and false alarm are necessary before the optimal decision fusion is conducted.

2.3.2 M-ary Hypothesis

After deriving the optimal fusion rule for binary distributed detection, we then move to the system with multiple hypotheses. For the jth local detector, we define the probability of anomaly as

$$\varepsilon_{ik}^j \triangleq P(u_j = H_k | H_i \text{ is true}) \tag{2.45}$$

where u_j is the decision of the jth local detector, $i, k \in \{0, \cdots, M-1\}$ and $i \neq k$. The anomaly probabilities measure performance of local detectors. Clearly,

$$\begin{aligned} \varepsilon_{ii}^j &\triangleq P(u_j = H_i | H_i \text{ is true}) \\ &= 1 - \sum_{\substack{k=0 \\ k \neq i}}^{M-1} \varepsilon_{ik}^j \end{aligned} \tag{2.46}$$

In a compact form, we define the performance matrix of the jth local detector as

$$\Upsilon_j = [\varepsilon_{ik}^j] \tag{2.47}$$

We further partition the indices of the local detectors into

$$\begin{aligned} \mathcal{S}_0 &\triangleq \{j|\ u_j = H_0,\ \forall j = 1, \cdots, N\} \\ &\vdots \\ \mathcal{S}_{M-1} &\triangleq \{j|\ u_j = H_{M-1},\ \forall j = 1, \cdots, N\} \end{aligned} \tag{2.48}$$

The fusion rule can be derived by minimizing the probability of error at the fusion center

$$P_e = \sum_{i=0}^{M-1} \sum_{\substack{k=0 \\ k \neq i}}^{M-1} P(H_i) P(u_0 = H_k | H_i \text{ is true}) \tag{2.49}$$

It has been shown in [17] that minimizing the error probability in Eq. (2.49) reduces to maximizing the posterior probability

$$\begin{aligned} P(H_i|\boldsymbol{u}) &= P(H_i|u_1, u_2, \cdots, u_N) \\ &= \frac{P(H_i)}{P(\boldsymbol{u})} P(u_1|H_i) \cdots P(u_N|H_i) \end{aligned} \tag{2.50}$$

where again $\boldsymbol{u} = (u_1, \cdots, u_N)$. For $i = 0, \cdots, M-1$, the global decision is therefore [7]

$$\begin{aligned} u_0 &= \arg\max_{H_i} P(H_i|\boldsymbol{u}) \\ &= \arg\max_{H_i} P(H_i) \prod_{j \in \mathcal{S}_0} \varepsilon_{i0}^j \cdots \prod_{j \in \mathcal{S}_{M-1}} \varepsilon_{iM-1}^j \end{aligned} \tag{2.51}$$

Again, as we mentioned, in order to achieve the optimal overall detection, the prior probabilities and the probabilities of anomaly defined in Eq. (2.45) have to be known.

2.4 Blind Algorithm for Binary Distributed Detection

If there are only two hypotheses involved, a blind adaptive algorithm has been proposed to estimate the unknown parameters [6]. For the moment, we suppose the

distributed detection system is composed of three local detectors. Let P_{ijk} denote the probability of $u_1 = i$, $u_2 = j$ and $u_3 = k$, where $i, j, k = 0$ or 1 depending on the local decision, and $P(H_1) = P_1$. According to Bayes rule and conditional independence of local decisions, we have

$$\begin{aligned} P_{ijk} &= P(u_1 = i|H_1)P(u_2 = j|H_1)P(u_3 = k|H_1)P_1 \\ &= P(u_1 = i|H_0)P(u_2 = j|H_0)P(u_3 = k|H_0)(1 - P_1) \end{aligned} \tag{2.52}$$

Note that

$$P(u_j = i|H_1) = \begin{cases} 1 - P_j^m, & \text{if } i = 1 \\ P_j^m, & \text{otherwise} \end{cases} \tag{2.53}$$

and

$$P(u_j = i|H_0) = \begin{cases} P_j^f, & \text{if } i = 1 \\ 1 - P_j^f, & \text{otherwise} \end{cases} \tag{2.54}$$

where P_j^m and P_j^f represent the probability of miss and false alarm of the jth local detector, respectively. Substituting Eq. (2.53) and (2.54) into Eq. (2.52) yields a set of nonlinear equations

$$\begin{cases} P_{000} = P_1^m P_2^m P_3^m P_1 + (1 - P_1^f)(1 - P_2^f)(1 - P_3^f)(1 - P_1) \\ P_{001} = P_1^m P_2^m (1 - P_3^m) P_1 + (1 - P_1^f)(1 - P_2^f) P_3^f (1 - P_1) \\ P_{010} = P_1^m (1 - P_2^m) P_3^m P_1 + (1 - P_1^f) P_2^f (1 - P_3^f)(1 - P_1) \\ P_{100} = (1 - P_1^m) P_2^m P_3^m P_1 + P_1^f (1 - P_2^f)(1 - P_3^f)(1 - P_1) \\ P_{111} = (1 - P_1^m)(1 - P_2^m)(1 - P_3^m) P_1 + P_1^f P_2^f P_3^f (1 - P_1) \\ P_{011} = P_1^m (1 - P_2^m)(1 - P_3^m) P_1 + (1 - P_1^f) P_2^f P_3^f (1 - P_1) \\ P_{110} = (1 - P_1^m)(1 - P_2^m) P_3^m P_1 + P_1^f P_2^f (1 - P_3^f)(1 - P_1) \\ P_{101} = (1 - P_1^m) P_2^m (1 - P_3^m) P_1 + P_1^f (1 - P_2^f) P_3^f (1 - P_1) \end{cases} \tag{2.55}$$

Here, we assume the probabilities P_{ijk} are known. Since the summation of all probabilities of the LHS of Eq. (2.55) is unity, only seven of them are independent. There are in total seven unknown parameters, therefore in principle they can be obtained by

solving the set of nonlinear equations. The complicated analytical solution is given in [6]. Define

$$X = \frac{\gamma_* - \gamma_1\gamma_2\gamma_3 - (\gamma_1 a_2 a_3 + \gamma_2 a_1 a_3 + \gamma_3 a_1 a_2)}{\sqrt{(\delta_{12} - \gamma_1\gamma_2)(\delta_{13} - \gamma_1\gamma_3)(\delta_{23} - \gamma_2\gamma_3)}} \tag{2.56}$$

where

$$\begin{aligned} \gamma_1 &= \sum_{j,k} P_{1jk} \\ \gamma_2 &= \sum_{i,k} P_{i1k} \\ \gamma_3 &= \sum_{i,j} P_{ij1} \\ \gamma_* &= P_{111} \end{aligned} \tag{2.57}$$

$$\begin{aligned} \delta_{12} &= P_{110} + P_{111} \\ \delta_{13} &= P_{101} + P_{111} \\ \delta_{23} &= P_{011} + P_{111} \end{aligned} \tag{2.58}$$

and

$$\begin{aligned} a_1 &= \sqrt{\frac{(\delta_{12} - \gamma_1\gamma_2)(\delta_{13} - \gamma_1\gamma_3)}{\delta_{23} - \gamma_2\gamma_3}} \\ a_2 &= \sqrt{\frac{(\delta_{12} - \gamma_1\gamma_2)(\delta_{23} - \gamma_2\gamma_3)}{\delta_{13} - \gamma_1\gamma_3}} \\ a_3 &= \sqrt{\frac{(\delta_{13} - \gamma_1\gamma_3)(\delta_{23} - \gamma_2\gamma_3)}{\delta_{12} - \gamma_1\gamma_2}} \end{aligned} \tag{2.59}$$

Therefore, the unknown parameters can be computed as

$$P_1 = 0.5 - \frac{X}{2\sqrt{X^2 + 4}} \tag{2.60}$$

and

$$\begin{aligned} P_j^f &= \gamma_j - a_j\sqrt{\frac{P_1}{1 - P_1}} \\ P_j^m &= 1 - \gamma_j - a_j\sqrt{\frac{1 - P_1}{P_1}} \end{aligned} \tag{2.61}$$

where $j = 1, 2, 3$. In practice, the probabilities defined in Eq. (2.55) are unavailable and replaced by empirical probabilities. As a result, the estimates of unknown parameters are obtained by solving Eq. (2.55) and asymptotically unbiased under some general conditions. This algorithm can be modified to accommodate to the distributed detection system with more than three local detectors. This extension is based on the fact the probability of false alarm and miss of one local detector is related to those of the other as follows:

$$P_j^f = \frac{\frac{P_1\delta_{ij} - \gamma_i\gamma_j}{1-P_1} + \gamma_j P_i^f}{P_i^f - \gamma_i} \tag{2.62}$$

and

$$P_j^m = 1 + \frac{\frac{\gamma_i\gamma_j - (1-P_1)\delta_{ij}}{P_1} - \gamma_j(1 - P_i^m)}{1 - \gamma_i - P_i^m} \tag{2.63}$$

where $\gamma_i = P(u_i = 1)$ and $\delta_{ij} = P(u_i = 1, u_j = 1)$, $i, j = 1, \cdots, N$ and $i \neq j$. We first arbitrarily select any three local detectors and use the blind algorithm in this section to estimate P_1 and their individual probability of false alarm and miss. Using the relationship described in Eq. (2.62) and (2.63), we then have the probabilities of false alarm and miss of other local detectors. Same to the case of three local detectors, the probabilities γ_i and δ_{ij} are not available in practice and are estimated using empirical probabilities.

This approach works well for binary distributed detection. Unfortunately, it is not extendable when multiple hypotheses are involved because analytical solution does not exist any more.

Chapter 3

Parameter Estimation

In practice, we have no knowledge on prior probabilities of the hypotheses and the local detector performance. We have, in general, only a set of decision sequences made by N local detectors. The blind algorithm proposed for binary distributed detection in [6] cannot be extended to M-ary case. In this chapter, we formulate the problem of estimating the unknown parameters and present two algorithms to estimate them using local decisions in order to achieve optimal M-ary decision fusion. We conclude this chapter with the derivation of the biased Cramér-Rao bound, which is a measure of the estimation efficiency. The estimation processes are only dependent on the local decisions, as a result, the corresponding fusion rule is usually referred to as blind fusion rule.

3.1 Problem Formulation

Let N_d be the total number of decisions made by a particular local detector. The set of decision sequences denoted by $\{u_{jn}\}$, $j = 1, \cdots, N$ and $n = 1, \cdots, N_d$ is explained as follows: for a fixed n, the set represents the decisions on the same phenomenon made by N local detectors. It represents the decision sequence of different phenomena made by a particular local detector when j is fixed. For M hypotheses and N local detectors,

let $\boldsymbol{U}$ be the set consisting of all of the possible decision combinations. Clearly, $\dim(\boldsymbol{U}) = M^N$. Let the random variable X_ℓ indicate the number of times the ℓth combination $\boldsymbol{u}_\ell$ occurring with occurrence probability $P(\boldsymbol{u}_\ell)$, where $\ell = 1, \cdots, L$ and $L = M^N$. We refer to X_ℓ as occurrence number. Furthermore, recall the Bayes' rule and conditional independence, the probability of one of the M^N possible combinations can be written as

$$\begin{aligned} p_\ell &= P(\boldsymbol{u}_\ell) \\ &= \sum_{i=0}^{M-1} P(H_i)P(u_1 = s_1, \cdots, u_N = s_N|H_i) \\ &= \sum_{i=0}^{M-1} P(H_i)P(u_1 = s_1|H_i)\cdots P(u_N = s_N|H_i) \end{aligned} \tag{3.1}$$

where $s_1, \cdots, s_N \in \{H_0, \cdots, H_{M-1}\}$. Substituting the true values of prior and anomaly probabilities into Eq. (3.1) gives all occurrence probabilities. For a fixed total number of local decisions N_d, the occurrence numbers of all possible decision combinations, namely $\boldsymbol{X} = (X_1, X_2, \cdots, X_L)$, are multinomially distributed with probability mass function

$$P(X_1 = x_1, \cdots, X_L = x_L|N_d) = \frac{N_d!}{x_1! \cdots x_L!} p_1^{x_1} \cdots p_L^{x_L} \tag{3.2}$$

and $\text{var}(X_\ell) = N_d p_\ell(1 - p_\ell)$, $\text{cov}(X_s X_\ell) = -N_d p_s p_\ell$ for $s = 1, \cdots, L$ and $s \neq \ell$ [26]. We define the vector $\boldsymbol{\varepsilon}$ consisting of $(M-1)MN$ unknown probabilities of anomaly in Eq. (2.45) and the $(MN+1)(M-1)$ – dimensional vector $\boldsymbol{\theta} = [\boldsymbol{\varepsilon}, P(H_0), \cdots, P(H_{M-2})]$. As illustrated in Eq. (3.1), the occurrence probability p_ℓ is the nonlinear function of unknown parameters represented by $\boldsymbol{\theta}$, i.e., $p_\ell = f_\ell(\boldsymbol{\theta})$.

3.2 Least Squares Estimation

For binary distributed detection, the unknown parameters can be obtained by analytically solving a set of nonlinear equations described in Eq. (3.1). It has also

been shown in [6] that the estimates of unknown parameters converge to their true values asymptotically. As for its M-ary counterpart, the set of nonlinear equations is not solvable directly. We therefore apply LS algorithm to estimate the unknown parameters. The occurrence probability is estimated as time averaging, i.e., empirical probability

$$\begin{aligned} p_\ell &= P(u_1 = s_1, u_2 = s_2, \cdots, u_N = s_N) \\ &\backsimeq \frac{\text{number of } (u_1 = s_1, \cdots, u_N = s_N)}{\text{total number of local decisions } N_d} \end{aligned} \tag{3.3}$$

where $s_1, \cdots, s_N \in \{H_0, \cdots, H_{M-1}\}$. Let y_ℓ be the estimate of the ℓth occurrence probability and recall the true occurrence probability $p_\ell = f_\ell(\boldsymbol{\theta})$, hence

$$y_\ell = f_\ell(\boldsymbol{\theta}) + e_\ell, \quad \ell = 1, \cdots, L \tag{3.4}$$

where e_ℓ is the estimation error. We define the vector $\boldsymbol{y} = [y_1, y_2, \cdots, y_L]^T$, $\boldsymbol{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \cdots, f_L(\boldsymbol{\theta})]^T$, and $\boldsymbol{e} = [e_1, e_2, \cdots, e_L]^T$, the problem can therefore be formulated as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{e} \tag{3.5}$$

Note that in the above model, there are, in general, M^N of such equations altogether of which only $M^N - 1$ are independent. These, combined with the constraining equation that the sum of all the occurrence probabilities is unity yield the set of equations for the parameters to be calculated. There are $(MN + 1)(M - 1)$ unknown parameters. We then utilize nonlinear least squares fitting to minimize $\| \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\theta}) \|^2$ [22].

3.3 Maximum Likelihood Estimation

To account for the known distribution of local decision combinations, the ML estimation is a very efficient alternative to apply. The principle of ML parameter estimation is to find the parameter values that make the observed data most likely. It is intuitively appealing and has remained one of the most powerful methods in estimation

theory, provided that the joint probability distribution of the available observed data set is formulated as the function of parameters of interest. The detailed discussion can be found in [17, 18]. As discussed before, the occurrence numbers of all possible local decision combinations X_ℓ are multinomially distributed with likelihood function

$$\begin{aligned} & P(X_1 = x_1, \cdots, X_L = x_L | N_d, \boldsymbol{\theta}) \\ = & \frac{N_d!}{x_1! \cdots x_L!} p_1^{x_1} \cdots p_L^{x_L} \end{aligned} \tag{3.6}$$

where $p_\ell = f_\ell(\boldsymbol{\theta})$, $\ell = 1, \cdots, L$. The essence of ML method is that we should estimate the parameter $\boldsymbol{\theta}$ by its most plausible values, given the numbers of all possible local decision combinations. In other words, once the occurrence numbers are known, the ML estimator is the value of $\boldsymbol{\theta}$ such that the joint likelihood function described in Eq. (3.6) is maximized, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} P(X_1 = x_1, \cdots, X_L = x_L | N_d, \boldsymbol{\theta}) \tag{3.7}$$

3.4 The Biased Cramér-Rao Bounds of Unknown Parameters

To examine how good the estimate is, we need to compare it with a benchmark quantifying the best estimate we may achieve under certain assumptions [17]. The Cramér-Rao bound (CRB) is a lower bound of the variance of any unbiased estimator. It is well known that in the case of the observations are related to the unknown parameters through a linear Gaussian model, the ML estimate achieves the CRB. However, in general, the estimates using a finite number of observations through linear and nonlinear models (such as the present case of the ML estimates of unknown parameters through a multinomial model) are biased [23]. Their variance is no longer bounded by the CRB, but by the biased CRB for a given bias gradient [17]. In the case where the choice of such bias gradient is not obvious, a uniform CRB for a scalar function of a deterministic vector parameter has been proposed in [24] to quantify the

smallest attainable variance using any estimator whose bias gradient norm is less than or equal to a constant. The results have been extended to the problem of estimating the uniform CRB for a parameter vector with bias gradient matrix norm bounded by a constant in [25]. In this section, we assume the bias gradient matrix is known and derive the biased CRB for unknown parameters. If $\hat{\boldsymbol{\theta}}$ is an estimate with bias vector $\boldsymbol{b}(\boldsymbol{\theta}_0)$, then the biased CRB is given by [17]

$$\boldsymbol{C}_{\hat{\theta}} \succeq (\boldsymbol{I} + \boldsymbol{B}(\boldsymbol{\theta}_0))\boldsymbol{J}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{I} + \boldsymbol{B}(\boldsymbol{\theta}_0))^T \tag{3.8}$$

where $\boldsymbol{C}_{\hat{\theta}}$ is the covariance matrix of the estimate, $\boldsymbol{\theta}_0$ is the true value of unknown parameter vector and assumed to be known for the purpose of deriving the biased CRB, $\boldsymbol{J}(\boldsymbol{\theta}_0)$ is the Fisher information matrix and $\boldsymbol{B}(\boldsymbol{\theta}_0) = \frac{\partial \boldsymbol{b}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is the known bias gradient matrix. As discussed earlier, there are L occurrence probabilities. Since the summation of all occurrence probabilities is unity, any one of them can be uniquely determined by the other $L-1$ ones. We let $\boldsymbol{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \cdots, f_{L-1}(\boldsymbol{\theta})]^T$ and $g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))$ be the nonlinear distribution function defined in Eq. (3.2). Therefore

$$\begin{aligned}
\boldsymbol{J}(\boldsymbol{\theta}) &= E_{\boldsymbol{X}}\left[\frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right]\left[\frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right]^T \\
&= E_{\boldsymbol{X}}\left\{\left[\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T \boldsymbol{g}(\boldsymbol{X})\boldsymbol{g}^T(\boldsymbol{X})\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} \\
&= \left[\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T E_{\boldsymbol{X}}\left\{\boldsymbol{g}(\boldsymbol{X})\boldsymbol{g}^T(\boldsymbol{X})\right\}\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}
\end{aligned} \tag{3.9}$$

where

$$\boldsymbol{g}(\boldsymbol{X}) = \frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial \boldsymbol{f}(\boldsymbol{\theta})} \tag{3.10}$$

Let us define the $(L-1) \times Q$ matrix $\boldsymbol{H}(\boldsymbol{\theta})$ to be $\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, where $Q = (MN+1)(M-1)$, therefore its qth column $\boldsymbol{h}_q$ is

$$\boldsymbol{h}_q = \frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \theta_q}, \quad q = 1, \cdots, Q \tag{3.11}$$

Next we will determine the the expectation part of Eq. (3.9). Let

$$\boldsymbol{S}(\boldsymbol{\theta}) = E_{\boldsymbol{X}}\left\{\boldsymbol{g}(\boldsymbol{X})\boldsymbol{g}^T(\boldsymbol{X})\right\} \tag{3.12}$$

Its ijth element is

$$\begin{aligned} S_{ij} &\triangleq E\left\{\left[\frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial f_i(\boldsymbol{\theta})}\right]\left[\frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial f_j(\boldsymbol{\theta})}\right]\right\} \\ &= -E\left\{\frac{\partial^2 \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial f_i(\boldsymbol{\theta}) \partial f_j(\boldsymbol{\theta})}\right\} \end{aligned} \tag{3.13}$$

If $i = j$, then

$$\begin{aligned} S_{ii} &= -E\left\{\frac{\partial^2 \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial^2 f_i(\boldsymbol{\theta})}\right\} \\ &= E\left\{\frac{X_i}{f_i^2(\boldsymbol{\theta})} + \frac{X_L}{f_L^2(\boldsymbol{\theta})}\right\} \end{aligned} \tag{3.14}$$

Otherwise,

$$S_{ij} = E\left\{\frac{X_L}{f_L^2(\boldsymbol{\theta})}\right\} \tag{3.15}$$

From [26], we know the mean of $X_i,\ i = 1, \cdots, L$ is

$$E(X_i) = N_d f_i(\boldsymbol{\theta}) \tag{3.16}$$

and apply it to Eq. (3.14) and (3.15), we have

$$S_{ij} = \begin{cases} N_d\left(\frac{1}{f_i(\boldsymbol{\theta})} + \frac{1}{f_L(\boldsymbol{\theta})}\right), & \text{if } i = j \\ \frac{N_d}{f_L(\boldsymbol{\theta})}, & \text{otherwise} \end{cases} \tag{3.17}$$

Consequently,

$$\boldsymbol{S}(\boldsymbol{\theta}) = N_d \boldsymbol{R}(\boldsymbol{\theta}) \tag{3.18}$$

where the $(L-1) \times (L-1)$ matrix $\boldsymbol{R}(\boldsymbol{\theta})$ is

$$\boldsymbol{R}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{f_1(\boldsymbol{\theta})} + \frac{1}{f_L(\boldsymbol{\theta})} & \cdots & \frac{1}{f_L(\boldsymbol{\theta})} \\ \vdots & \ddots & \vdots \\ \frac{1}{f_L(\boldsymbol{\theta})} & \cdots & \frac{1}{f_{L-1}(\boldsymbol{\theta})} + \frac{1}{f_L(\boldsymbol{\theta})} \end{bmatrix} \tag{3.19}$$

As a result,

$$\boldsymbol{J}(\boldsymbol{\theta}) = N_d \boldsymbol{H}^T(\boldsymbol{\theta}) \boldsymbol{R}(\boldsymbol{\theta}) \boldsymbol{H}(\boldsymbol{\theta}) \tag{3.20}$$

The Fisher information matrix can be obtained by substituting $\boldsymbol{\theta}_0$ into Eq. (3.20) and the bound of estimate is computed accordingly.

3.5 Number of Unknown Parameters

The number of unknown parameters to be estimated increases exponentially with the number of local detectors. When there are a large number of local detectors, the computational complexity of estimating all of the unknown parameters could be prohibitively high. However, depending on the model of the phenomenon, the number of unknown parameters may be reduced in some cases. To illustrate more clearly, let us consider a binary hypothesis testing problem in wireless sensor networks. The observation at the jth sensor (local detector) is

$$\begin{aligned} H_1: &\quad z_j = s_j + n_j \\ H_0: &\quad z_j = n_j \end{aligned} \tag{3.21}$$

where s_j is the signal amplitude. The signal power at the sensor is generally a decreasing function of the signal power at the signal source location P_o and the distance between the sensor and source location d_j, i.e.,

$$s_j^2 = \mu(P_o, d_j) \tag{3.22}$$

for example, in free space the power density of all electromagnetic waves such as radio, light and X-rays is inversely proportional to the square of the distance from the source [27]. We assume that the noises at different sensors are i.i.d and follow Gaussian distribution $\mathcal{N}(0, \sigma^2)$. All sensors perform LRT to make local decisions using the identical threshold τ. Therefore, the false alarm probabilities of all sensors are the same and defined as

$$P_j^f = \int_{\tau}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx \tag{3.23}$$

However, the probabilities of miss defined as

$$\begin{aligned} P_j^m &= \int_{-\infty}^{\tau} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-s_j)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\tau} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\sqrt{\mu(P_o,d_j)})^2}{2\sigma^2}} dx \end{aligned} \tag{3.24}$$

are dependent on the sensor location. In this case, there are totally $N + 2$ unknown parameters rather than $2^N + 1$. If we know that all of the distances d_j are functions of a small number of unknown parameters, the total number of parameters to be estimated can be further reduced.

Chapter 4

Overall Detection Performance

In this chapter, we derive analytically the optimal detection performance in the sense of overall probability of error when the prior probabilities and the parameters describing the local detector performance are known and have a discussion about the case when they are unknown and estimated using our blind algorithm.

4.1 Optimal Detection Performance

We assume in this section that the prior probabilities and parameters quantifying the local detection performance are known.

4.1.1 Binary Distributed Detection

For binary distributed detection, the fusion center makes a binary global decision based on the following fusion rule [6]

$$w_0 + \sum_{j=1}^{N} w_j \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{4.1}$$

where $w_0 = \log \frac{P_1}{P_0}$ and

$$w_j = \begin{cases} \log \frac{1-P_j^m}{P_j^f}, & \text{if } u_j = H_1 \\ \log \frac{P_j^m}{1-P_j^f}, & \text{otherwise} \end{cases} \tag{4.2}$$

where P_j^f and P_j^m are probability of false alarm and miss of the jth local detector, respectively, and $j = 1, \cdots, N$. When H_0 is true, the weight w_j is a random variable with the following distribution

$$w_j|H_0 = \begin{cases} \log \frac{1-P_j^m}{P_j^f}, & \text{with prob } P_j^f \\ \log \frac{P_j^m}{1-P_j^f}, & \text{with prob } 1 - P_j^f \end{cases} \tag{4.3}$$

Similarly, when H_1 is true,

$$w_j|H_1 = \begin{cases} \log \frac{1-P_j^m}{P_j^f}, & \text{with prob } 1 - P_j^m \\ \log \frac{P_j^m}{1-P_j^f}, & \text{with prob } P_j^m \end{cases} \tag{4.4}$$

The sum of the weights $w_0 + \sum_{j=1}^{N} w_j$ can take on 2^N different possible values depending on the decisions made by local detectors. If the local decisions are $(s_1, \cdots, s_N)$, where $s_1, \cdots, s_N \in \{H_0, H_1\}$, the probability of the sum taking on the corresponding value is $P(u_1 = s_1|H_i) \cdots P(u_N = s_N|H_i)$. As a result, the overall probability of error is

$$\begin{aligned} P_e &= \sum_{i=0}^{1} \sum_{\substack{k=0 \\ k \neq i}}^{1} P(H_i) P(u_0 = H_k|H_i) \\ &= P(H_0) P\left[w_0 + \sum_{j=1}^{N} w_j > 0|H_0\right] + P(H_1) P\left[w_0 + \sum_{j=1}^{N} w_j < 0|H_1\right] \\ &= P(H_0) P\left[w_0 + \sum_{j=1}^{N} w_j|H_0 > 0\right] + P(H_1) P\left[w_0 + \sum_{j=1}^{N} w_j|H_1 < 0\right] \end{aligned} \tag{4.5}$$

4.1.2 M-ary Distributed Detection

In the M-ary case, $M-1$ likelihood ratios are necessary to derive the detection scheme. For $\ell = 0, \cdots, M-1$, let us define

$$\Lambda_\ell(\boldsymbol{u}) = \frac{P(\boldsymbol{u}|H_\ell)}{P(\boldsymbol{u}|H_0)} \tag{4.6}$$

and

$$\phi_\ell(\boldsymbol{u}) = \sum_{\substack{j=0 \\ j\neq\ell}}^{M-1} P(H_j)P(\boldsymbol{u}|H_j) \tag{4.7}$$

The global decision is the hypothesis corresponding to the minimum value of $\phi_\ell(\boldsymbol{u})$ as shown in [17]. Dividing Eq. (4.7) by $P(\boldsymbol{u}|H_0)$ yields

$$\varphi_\ell(\boldsymbol{u}) = \sum_{\substack{j=0 \\ j\neq\ell}}^{M-1} P(H_j)\Lambda_\ell(\boldsymbol{u}) \tag{4.8}$$

The global decision rule can be written as

$$u_0 = H_\ell \quad \text{if } \varphi_\ell(\boldsymbol{u}) = \min\{\varphi_0(\boldsymbol{u}), \cdots, \varphi_{M-1}(\boldsymbol{u})\} \tag{4.9}$$

The decision space is an $M-1$ dimensional space spanned by the likelihood ratios $\Lambda_1(\boldsymbol{u}), \cdots, \Lambda_{M-1}(\boldsymbol{u})$. We then define the weight for each likelihood ratio similar to that in binary detection. Let

$$w_0^\ell = \log\frac{P(H_\ell)}{P(H_0)}, \quad \forall\ \ell = 1, \cdots, M-1 \tag{4.10}$$

and for $\ell = 1, \cdots, M-1$, let w_j^ℓ be the weight of the ℓth likelihood ratio defined in Eq. (4.6). If the jth local detector makes a decision in favor of H_k, then

$$\begin{aligned} w_j^\ell &= \log\frac{P(u_j = H_k|H_\ell)}{P(u_j = H_k|H_0)} \\ &= \log\frac{\varepsilon_{\ell k}^j}{\varepsilon_{0k}^j}, \quad \forall\ j = 1, \cdots, N \end{aligned} \tag{4.11}$$

The weight w_j^ℓ is a random variable and can take on M possible values with known probabilities depending only on the decision made at the jth local detector, i.e.,

$$\begin{aligned} & P\left(w_j^\ell = \log\frac{\varepsilon_{\ell k}^j}{\varepsilon_{0k}^j} | H_i \text{ is true}\right) \\ = \; & P(u_j = H_k | H_i \text{ is true}) \\ = \; & \varepsilon_{ik}^j \end{aligned} \tag{4.12}$$

Once the decision at the jth local detector is determined, the values of all weights corresponding to the same detector w_j^ℓ, $\ell = 1, \cdots, M-1$ are known. Substituting all weights w_0^ℓ and w_j^ℓ into the global decision rule in Eq. (4.9) arrives at the global decision. Consequently, the global decision rule for M-ary distributed detection can be written in a compact form

$$u_0 = \begin{cases} H_0, & \text{if } \sum_{j=0}^N w_j^\ell < 0, \forall \ell = 1, \cdots, M-1 \\ H_k, & \text{if } \sum_{j=0}^N w_j^k > 0 \text{ and } \sum_{j=0}^N w_j^k = \\ & \quad \max\{\sum_{j=0}^N w_j^1, \cdots, \sum_{j=0}^N w_j^{M-1}\} \end{cases} \tag{4.13}$$

Therefore the probability of error is

$$\begin{aligned} P_e &= 1 - \sum_{i=0}^{M-1} P(H_i) P(u_0 = H_i | H_i) \\ &= 1 - P(H_0) P\left[\sum_{j=0}^N w_j^1 < 0, \cdots, \sum_{j=0}^N w_j^{M-1} < 0 | H_0\right] \\ &\quad - \sum_{i=1}^{M-1} P(H_i) P\left[\sum_{j=0}^N w_j^i > 0, E_i | H_i\right] \end{aligned} \tag{4.14}$$

where the event E_i is defined as

$$E_i = \sum_{j=0}^N w_j^i > \sum_{j=0}^N w_j^n, \quad \forall\, n = 1 \cdots, M-1 \text{ and } n \neq i \tag{4.15}$$

Eq. (4.5) and (4.14) clearly state that the overall error probability is the function of the prior probabilities and weights w_j or w_j^ℓ, and the weights are the functions of system parameters. If all of them are known, the error probability can be computed.

4.2 Detection Performance Using Blind Algorithm

In realistic scenarios, we have to estimate those unknown parameters using local decisions. The LS and ML estimates are continuous random variables and asymptotically unbiased under suitable regularity conditions [22]. Meanwhile, the estimation accuracy may also be enhanced if a more efficient estimation algorithm is available. These require longer processing time and have higher computational complexity. However, the purpose of distributed detection is to improve overall detection performance. If the estimation accuracy arrives at a level above which the overall detection performance cannot be improved significantly, it may be meaningless to make further effort on estimation accuracy. We then say the system is robust to the unknown parameters. This is reasonable because in essence the global decision is made by determining which decision region the local decisions fall into. Since the decision regions are dependent on the unknown parameters. The estimates of unknown parameters may not be precise, as long as they are be such that the decision regions remain unchanged, the overall error probability is the same. The overall probability of error is a continuous random variable varying from experiment to experiment. This follows from the fact that the estimates of unknown parameters are random. In some special cases, if the distribution function of both prior and anomaly probabilities are known, the distribution of overall error probability may be obtained accordingly because the error probability is the nonlinear function of unknown parameters [28]. However, in general the distribution function of error probability cannot be evaluated analytically and Monte Carlo simulations have to be performed. It should be noted that if the hypotheses are not exhaustive, the parameters describing the performance of local detectors may not be continuous random variables any more, but rather have mixture distributions, which will be taken into account in our future work.

The local decisions follow the probability distribution whose parameters are unknown. When we use their estimates to perform distributed detection, the resulting

overall detection performance is not identical to the one by substituting the estimates into Eq. (4.5). It is the system parameter mismatch that results in such discrepancy. In the following, we will analyze this effect by considering the case where we only have two equally likely hypotheses and two local detectors. The decision rule and definition of weights are the same as Eq. (4.1) and (4.2) except $w_0 = 0$. When H_0 is true,

$$\hat{w}_1 + \hat{w}_2|H_0 = \begin{cases} \log \frac{(1-\hat{P}_1^m)(1-\hat{P}_2^m)}{\hat{P}_1^f \hat{P}_2^f}, & \text{with prob } P_1^f P_2^f \\ \log \frac{(1-\hat{P}_1^m)\hat{P}_2^m}{\hat{P}_1^f(1-\hat{P}_2^f)}, & \text{with prob } P_1^f(1-P_2^f) \\ \log \frac{\hat{P}_1^m(1-\hat{P}_2^m)}{(1-\hat{P}_1^f)\hat{P}_2^f}, & \text{with prob } (1-P_1^f)P_2^f \\ \log \frac{\hat{P}_1^m \hat{P}_2^m}{(1-\hat{P}_1^f)(1-\hat{P}_2^f)}, & \text{with prob } (1-P_1^f)(1-P_2^f) \end{cases} \tag{4.16}$$

Similar result can be obtained for $\hat{w}_1 + \hat{w}_2|H_1$. The error probability is therefore

$$\hat{P}_e = \frac{1}{2}P(\hat{w}_1 + \hat{w}_2 > 0|H_0) + \frac{1}{2}P(\hat{w}_1 + \hat{w}_2 < 0|H_1) \tag{4.17}$$

If we substitute all of the estimates into Eq. (4.5), the true probabilities in Eq. (4.16) are replaced by the corresponding estimates, which leads to the discrepancy of overall error probabilities. Although we restrict ourselves to consider the special system configuration, the idea is applicable to any distributed detection system where the unknown parameters are estimated using our blind algorithm.

Chapter 5

Numerical Examples and Discussions

In this chapter, we numerically investigate the applicability of our results in previous chapters and evaluate their individual performance. In all examples, we assume the distributed detection system as shown in Fig. 2.1 consisting of three local detectors and consider a detection scenario with $M = 3$ hypotheses unless otherwise stated.

5.1 Parameter Estimation

We use the first example to present the applicability of LS and ML algorithms for blindly estimating the unknown parameters. We assume the three hypotheses are equally likely. Decisions from three local detectors are generated based on the following performance matrices defined in Eq. (2.47)

$$\Upsilon_1 = \begin{bmatrix} 0.87 & 0.07 & 0.06 \\ 0.09 & 0.83 & 0.08 \\ 0.15 & 0.11 & 0.74 \end{bmatrix} \tag{5.1}$$

$$\Upsilon_2 = \begin{bmatrix} 0.86 & 0.05 & 0.09 \\ 0.05 & 0.88 & 0.07 \\ 0.12 & 0.08 & 0.8 \end{bmatrix} \tag{5.2}$$

and

$$\Upsilon_3 = \begin{bmatrix} 0.84 & 0.08 & 0.08 \\ 0.12 & 0.78 & 0.1 \\ 0.17 & 0.13 & 0.7 \end{bmatrix} \tag{5.3}$$

There are totally 20 unknown parameters to be estimated. We then apply LS and ML algorithm to the local decisions. In Fig. 5.1 and 5.2, the estimation accuracy of the proposed algorithms are evaluated based on 10000 runs. The average estimate of one particular parameter evolving with the number of local decisions is shown in Fig. 5.1 and the MSE of both estimators is depicted in Fig. 5.2. As expected, the ML estimator outperforms the LS estimator. The MSE of ML estimator is relatively small and does not decrease significantly with the increasing number of local decisions. The biased CRB, together with the sample variance of LS and ML estimate of one particular unknown parameter are shown in Fig. 5.3. Since the biased gradient matrix of ML estimates necessary for computing the bound is unavailable in reality, we estimate it using the method proposed in [24]. Let us recall in Section 3.4 we used $g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))$ to represent the likelihood function of all local decision combinations. Suppose we run the simulation T times. The biased gradient matrix can be estimated by

$$\begin{aligned} \widehat{\boldsymbol{B}(\boldsymbol{\theta}_0)} &= \frac{1}{T-1}\sum_{i=1}^{T}\left(\hat{\boldsymbol{\theta}}_i - \frac{1}{T}\sum_{j=1}^{T}\hat{\boldsymbol{\theta}}_j\right)\frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} - \boldsymbol{I} \\ &= \frac{1}{T-1}\sum_{i=1}^{T}\left(\hat{\boldsymbol{\theta}}_i - \frac{1}{T}\sum_{j=1}^{T}\hat{\boldsymbol{\theta}}_j\right)\left[\frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T \frac{\partial \log g(\boldsymbol{X}, \boldsymbol{f}(\boldsymbol{\theta}))}{\partial \boldsymbol{f}(\boldsymbol{\theta})} - \boldsymbol{I} \end{aligned} \tag{5.4}$$

where $\hat{\boldsymbol{\theta}}_i$, $i = 1, \cdots, T$ represents the estimate of the ith run. As can be clearly seen, the variances of LS and ML estimate are both bounded by the biased CRB and ML estimator outperforms LS estimator because its variance is closer to the biased CRB.

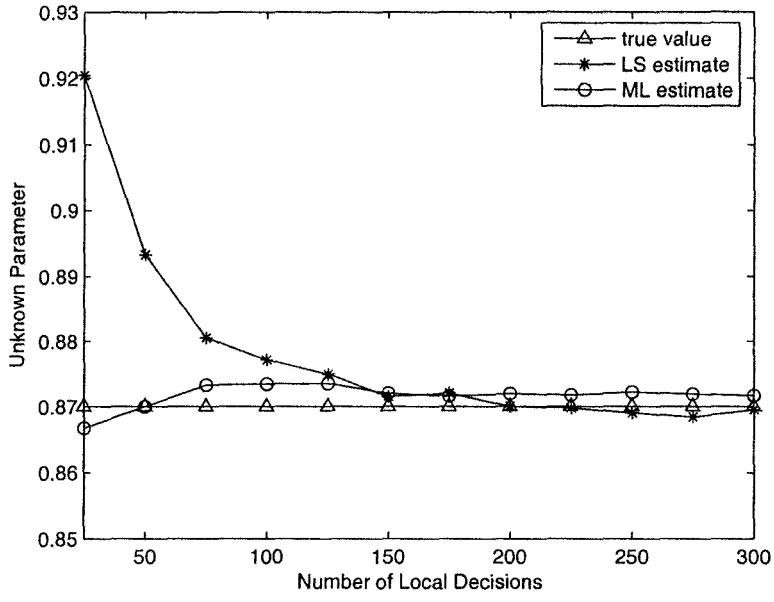

Figure 5.1: The evolution of LS and ML estimate with the local decision number.

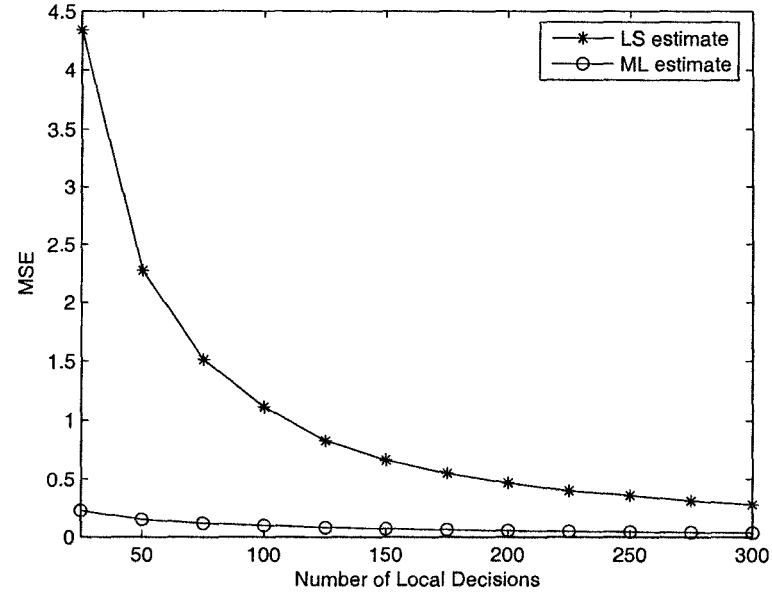

Figure 5.2: Comparison of MSE.

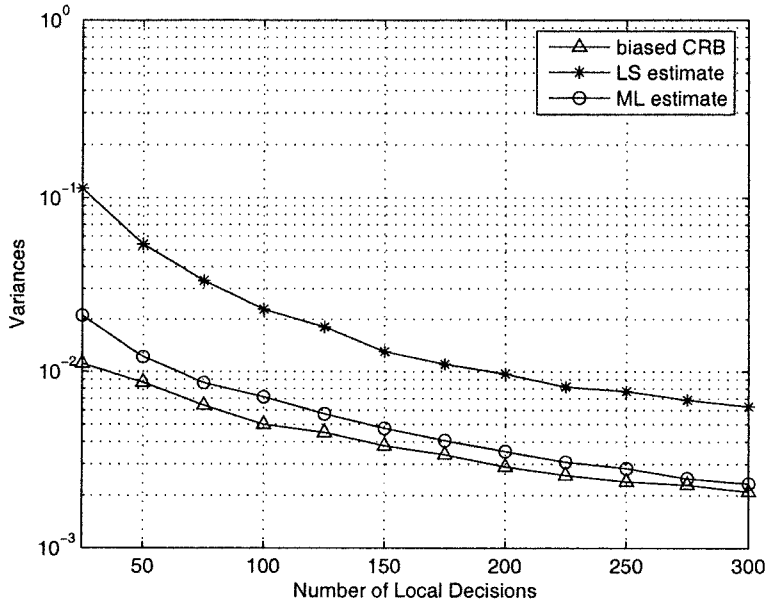

Figure 5.3: Variance comparison.

5.2 Overall Detection Performance

In this section, we use the LS and ML estimates obtained in the previous section as the system parameters necessary for distributed detection. In Fig. 5.4, we demonstrate how overall probability of error evolves with the increasing number of local decisions using two different estimation algorithms for 3-ary distributed detection system. The theoretical value is computed by substituting the true values of both prior and anomaly probabilities into Eq. (4.14). The other two overall error probability curves are obtained by averaging 10000 runs. In general, the overall error probability using ML estimates is smaller and approaches its true value faster than that using LS estimates. We can also see from the same figure that for a particular set of unknown parameters, there exists a threshold above which increasing number of local decisions will not yield significant improvements in the performance. In realistic scenarios, we cannot run the simulation many times to have the average estimates

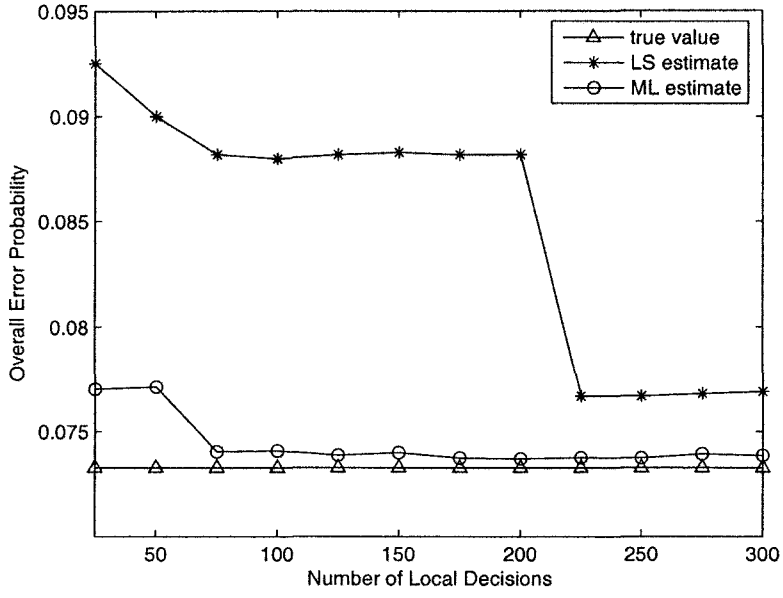

Figure 5.4: Overall performance of 3-ary distributed detection system.

of unknown parameters. To illustrate the overall detection performance using the estimates based on only one run, we show the histogram of overall error probability where the one-run estimates are obtained using different number of local decisions in Fig. 5.5. We have shown that Eq. (4.14) gives the theoretical value of overall error probability only if the true values of unknown parameters are known. When we just know their estimates, the overall detection performance can only be assessed through Monte Carlo simulation. However, direct substitution of the estimates into Eq. (4.14) gives a rough idea of the overall detection performance. The difference of overall error probabilities obtained by these two methods can be seen in Fig. 5.6. Its origin is clearly stated in Section 4.2.

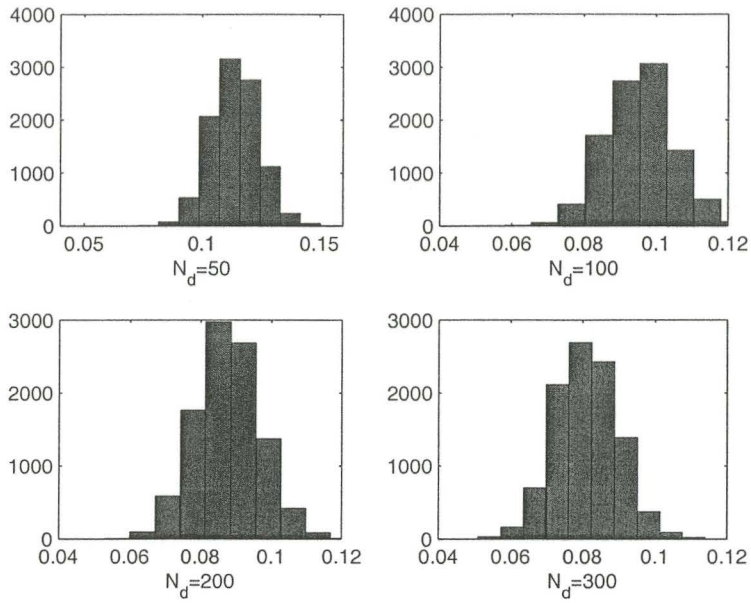

Figure 5.5: Histogram of overall error probability.

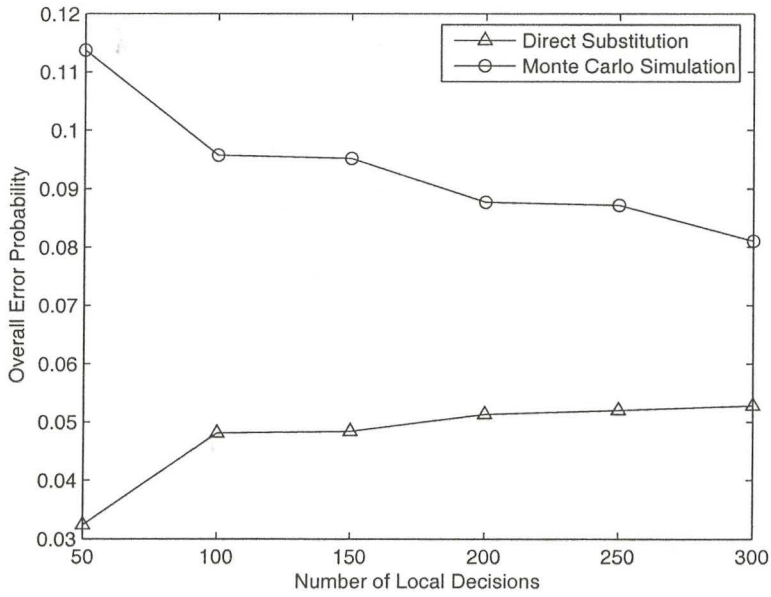

Figure 5.6: Comparison of overall error probability using direct substitution and Monte Carlo simulation.

5.3 Impact on Overall Detection Performance

In this section, we investigate the effect of several possible factors such as estimation error, number of local detectors and number of hypotheses to the overall detection performance. In Fig. 5.7, we assume there are only two local detectors. We want to investigate the impact on the overall error probability when the parameters we use (the estimated values) mismatch the true system parameters due to either small local decision number or low efficiency of estimation algorithm. For simplicity, we only consider the extreme case that all of the estimates deviate from their corresponding true values by the same percentage. From Fig. 5.7, the overall error probability exhibits piece-wise constant behavior with respect to the estimation error, showing that there are thresholds at which the error probability jumps to a new level. The piece-wise constant behavior is due to the fact that although the estimated values of the weights defined in Eq. (4.10) and (4.11) may change resulting from different estimation error, it is likely that the decision regions obtained from them remain unchanged, hence the overall error probability is constant. Since in realistic scenarios, none of the thresholds are known, the estimation error should be kept as small as possible.

In Fig. 5.8, we evaluate the overall performance by adding a new local detector. For equally likely binary hypotheses, we arbitrarily choose two local detectors with $P_1^m = 0.14$, $P_1^f = 0.1$, $P_2^m = 0.05$ and $P_2^f = 0.15$. From Eq. (4.5), we can compute the overall error probability to be 0.1. The overall error probability generally reduces if we add one more local detector. However, when the new local detector performs very poorly, the corresponding weight defined in Eq. (4.2) is very small. As a result, the system tends to ignore the new local detector. As for Fig. 5.9, there are only two equally likely hypotheses and one local detector in the beginning with arbitrarily selected performance parameters $P_1^m = 0.2$ and $P_1^f = 0.1$. For simplicity, we assume the new local detectors added are identical and the probability of miss and false alarm are the same. As expected, we can see the more local detectors with good

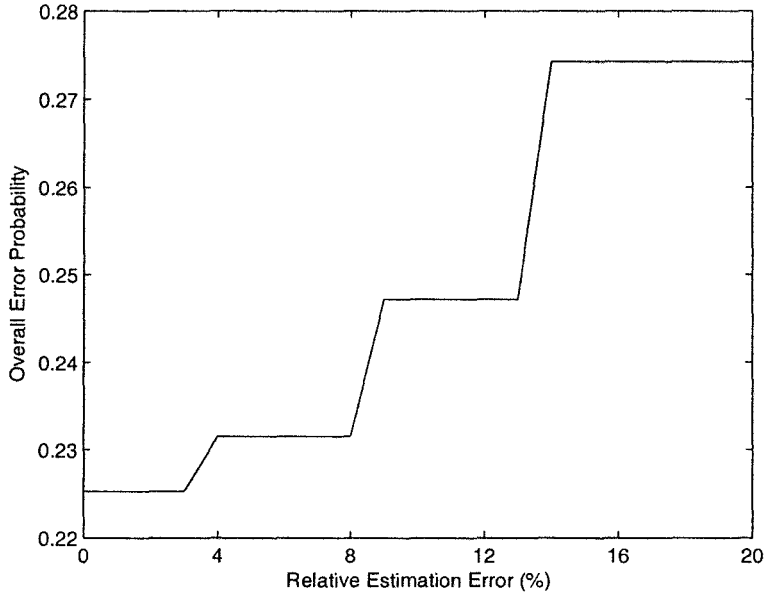

Figure 5.7: Impact of parameter mismatch on overall performance.

performance, the better the overall detection performance.

In Fig. 5.10, we compare the overall error probability in two cases: (a) the prior probabilities and probabilities of anomaly are known for different M, and (b) they are unknown and estimated using our blind algorithm with ML estimation. As for the case with 4 hypotheses, we assume that all of the local detectors are identical, i.e., they have the same probabilities of anomaly, to reduce the computational complexity. All simulations are conducted using 300 local decisions. The particular number of local decisions is picked because the overall detection performance after 300 local decisions does not improve significantly. It can be seen clearly that the performance of the scenario with estimated system parameters is very close to that with known prior probabilities and probabilities of anomaly.

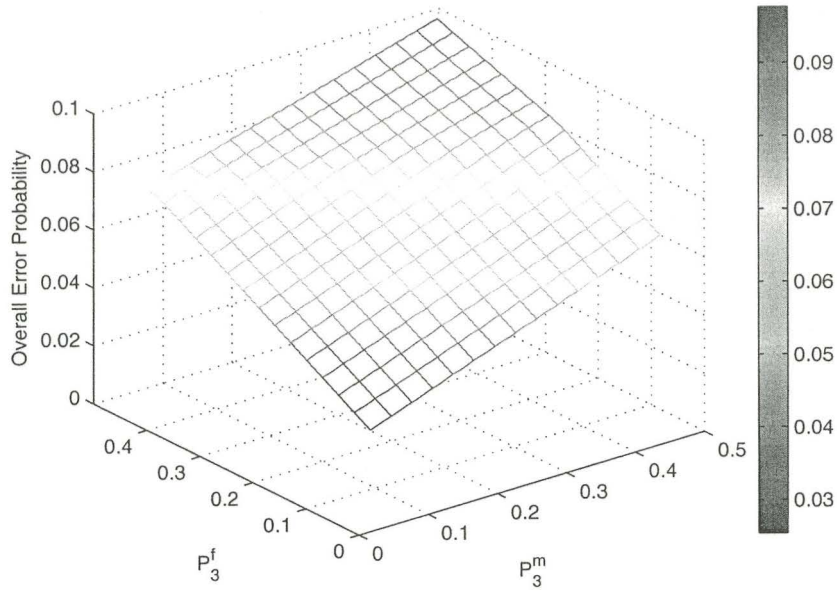

Figure 5.8: Overall performance with one more local detector.

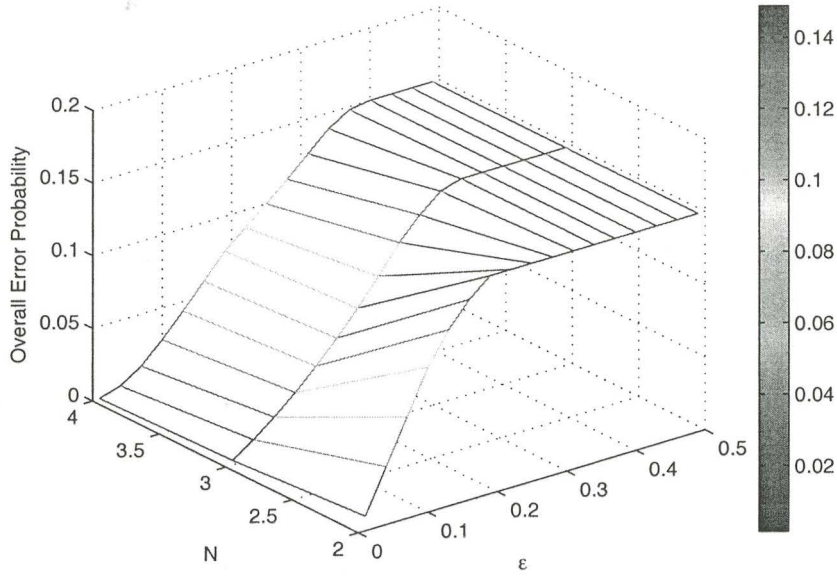

Figure 5.9: Impact of number of local detectors on overall performance.

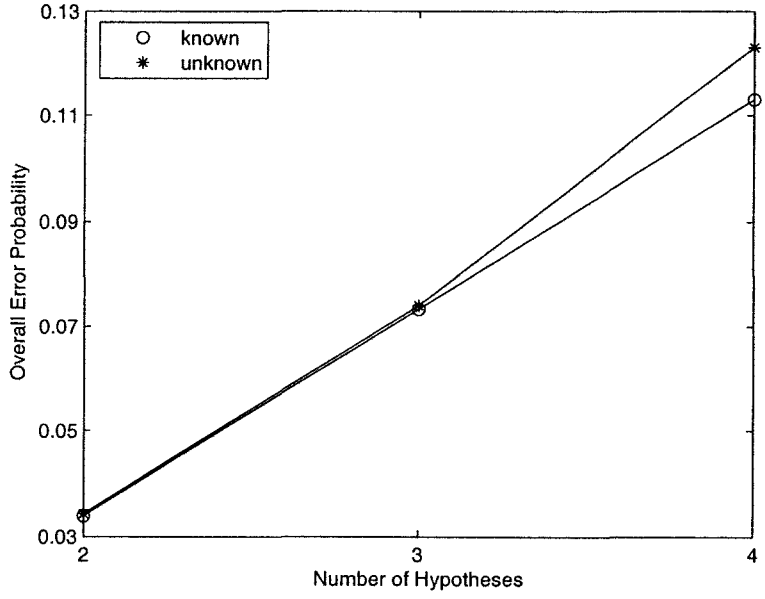

Figure 5.10: Impact of number of hypotheses on overall performance.

Chapter 6

Conclusion

In this thesis, we have investigated M-ary distributed detection problems under given local decision rules without assuming prior probabilities and the performance of local detectors are known at the fusion center. The previous research has been restricted to the system with known system parameters or binary distributed detection system. Providing the solution for M-ary distributed detection system and analytically analyzing the system performance are the main contributions of this thesis.

We started from the introduction of optimal fusion rule with fixed local decision rules. Then we used LS and ML algorithms to blindly estimate unknown parameters necessary for the optimal fusion. The biased CRBs of unknown parameters were also derived as a benchmark to compare the estimation performance. The ML estimator takes advantage of the parametric form of the distribution of local decisions and outperforms the LS estimator.

Further, we derived the analytical form of overall error probability and evaluated the detection performance for both binary and M-ary distributed detection. We also discussed the effect of the estimation inaccuracy, the number of local detectors and the number of hypotheses to the overall detection performance. The performance of

our proposed techniques were evaluated through numerical examples.

There are some interesting extensions of the research work. One of them is that we may drop the assumption of conditional independence of decisions at all local detectors. The local decisions for a specific hypothesis may be correlated from local detector to local detector or the decisions of specific local detectors are correlated from hypothesis to hypothesis. This phenomenon is quite typical in real life, such as distributed detection for biomedical signals. Another possible extension is to design the optimal fusion rule when the system parameters are not deterministic since in some cases they may change with environments. Finally, joint design of the local decision rules and fusion rule for correlated observations or time varying system parameters is a challenging but fruitful extension.

Bibliography

[1] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I– fundamentals", *Proceedings of IEEE*, vol. 85, pp. 54-63, Jan. 1997.

[2] R. S. Blum, S. A. Kassam and H. V. Poor, "Distributed detection with multiple sensors: Part II– advanced topics", *Proceedings of IEEE*, vol. 85, pp. 64-79, Jan. 1997.

[3] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aeros. and Elec. Sys.*, vol. AES-17, pp. 501-510, Jul. 1981.

[4] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems", *IEEE Trans. Aeros. and Elec. Sys.*, vol. 22, pp. 98-101, Jan. 1986.

[5] N. Ansari, E. S. H. Hou, B. Zhu and J. Chen, "Adaptive fusion by reinforcement learning for distributed detection systems", *IEEE Trans. Aeros. and Elec. Sys.*, vol. 32, pp. 524-531, Apr. 1996.

[6] G. Mirjalily, Z-Q Luo, T. D. Davidson and E. Bosse, "Blind adaptive decision fusion for distributed detection", *IEEE Trans. Aeros. and Elec. Sys.*, vol. 39, pp. 34-52, Jan. 2003.

[7] W. Baek and S. Bommareddy, "Optimal m-ary data fusion with distributed sensors", *IEEE Trans. Aeros. and Elec. Sys.*, vol. 31, pp. 1150-1152, July 1995.

[8] Q. Zhang and P. K. Varshney, "Decentralized M-ary detection via hierarchical binary decision fusion", *Inform. Fusion*, vol. 2, pp. 3-16, 2001.

[9] X. Zhu, Y. Yuan, C. Rorres and M. Kam, "Distributed M-ary hypothesis testing with binary local decisions", *Inf. Fusion*, vol. 5, pp. 157-167, 2004.

[10] B. Liu, A. Jeremic and K. M. Wong, "Blind adaptive algorithm for M-ary distributed detection", in *Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, Honolulu, Hawaii, Apr. 2007.

[11] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection", *IEEE Trans. Inform. Theory*, vol. 35, pp. 995-1000, Sept. 1989.

[12] B. Chen and P. K. Willett, "On the optimality of likelihood-ratio test for local sensor decision rules in the presence of nonideal channels", *IEEE Trans. Inform. Theory*, vol. 51, pp. 693-699, Feb. 2005.

[13] Y. Zhu, R. S. Blum, Z-Q. Luo and K. M. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases", *IEEE Trans. Autom. Control*, vol. 45, pp. 62-72, Jan. 2000.

[14] R. Srinivasan, "Distributed radar detection theory", *IEE Proceedings*, vol. 133, pp. 55-60, Feb. 1986.

[15] S. C. A. Thomopoulos, R. Viswanathan and D. C. Bougoulias, "Optimal decision fusion in multiple sensor systems", *IEEE Trans. Aeros. and Elec. Sys.*, vol. 23, pp. 644-653, Sept. 1987.

[16] A. Naim and M. Kam, "On-line estimation of probabilities for distributed Bayesian detection", *Automatica*, vol. 30, pp. 633-642, Apr. 1994.

[17] H. L. Van Trees, *Detection, Estimation, Modulation Theory, Part I.* New York: John Wiley & Sons, 1968.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volumn I: Estimation Theory*. New Jersey: Prentice-Hall, 1993.

[19] J. C. Mosher, R. M. Leahy and P. S. Lewis, "EEG and MEG: forward solutions for inverse methods", *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 245-259, Mar. 1999.

[20] J. C. Mosher, P. S. Lewis and R. M. Leahy, "Multiple dipole modelling and localization from spatio-temporal MEG data", *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 541-557, June 1992.

[21] A. Jeremic and A. Nehorai, "Design of chemical sensor arrays for monitoring disposal sites on the ocean floor", *IEEE Jour. Ocea. Eng.*, vol. 23, pp. 334-343, Oct. 1998.

[22] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York: John Wiley & Sons, 1989.

[23] R. A. Fisher, *Statistical Methods and Scientific Inference*, 3rd ed. New York: Hafner Press, 1973.

[24] A. O. Hero, J. A. Fessler and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound", *IEEE Trans. Signal Processing*, vol. 44, pp. 2026-2041, Aug. 1996.

[25] Y. C. Eldar, "Minimum variance in biased estimation: bounds and asymptotically optimal estimators", *IEEE Trans. Signal Processing*, vol. 52, pp. 1915-1930, July 2004.

[26] M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, 3rd ed. New York: John Wiley & Sons, Inc. 2000.

[27] L. Boithias, *Radio Wave Propagation*, New York: McGraw-Hill, 1987

[28] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.

[29] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, New Jersey: Prentice-Hall, 1974.